

# X-Net: Bimodal Feature Representation Learning for Satellite Images

Kenneth Tran, Department of Computer Science, NCSU

Wesam Sakla, Computational Engineering Division, LLNL

Hamid Krim, Department of Electrical and Computer Engineering, NCSU

# Primary Objective

Our aim is to create a meaningful joint representation across different modality data ( e.g. RGB and multispectral imagery) using deep learning with a goal of

- Identification of nuclear activity in satellite images
- Temporal analysis of satellite images to detect changes

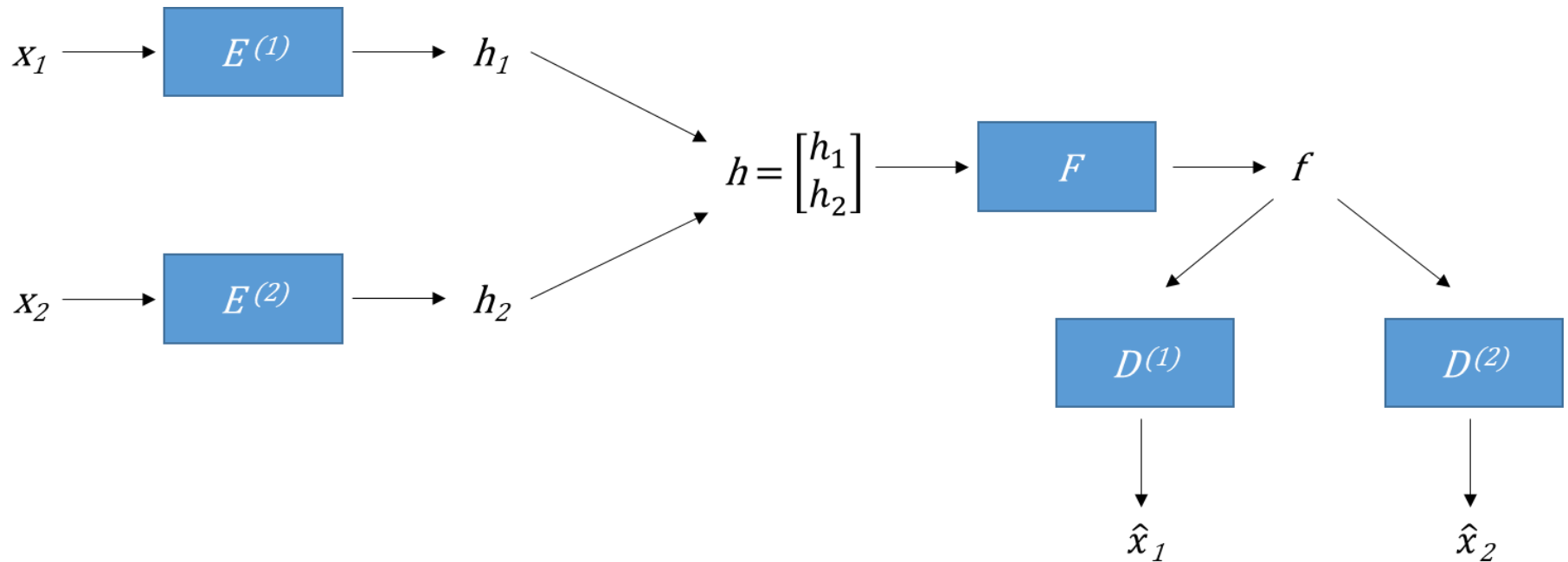
# Data Set: Functional Map of the World

- 1,047,691 images covering a majority of countries
- Includes 62 named categories and an additional false detection category
- Features metadata and statistics such as the ISO Country Code, UTM Zone, and Off-Nadir Angle
- With RGB and multispectral imagery content
- Each area of interest is imaged at multiple times, making temporal analysis possible

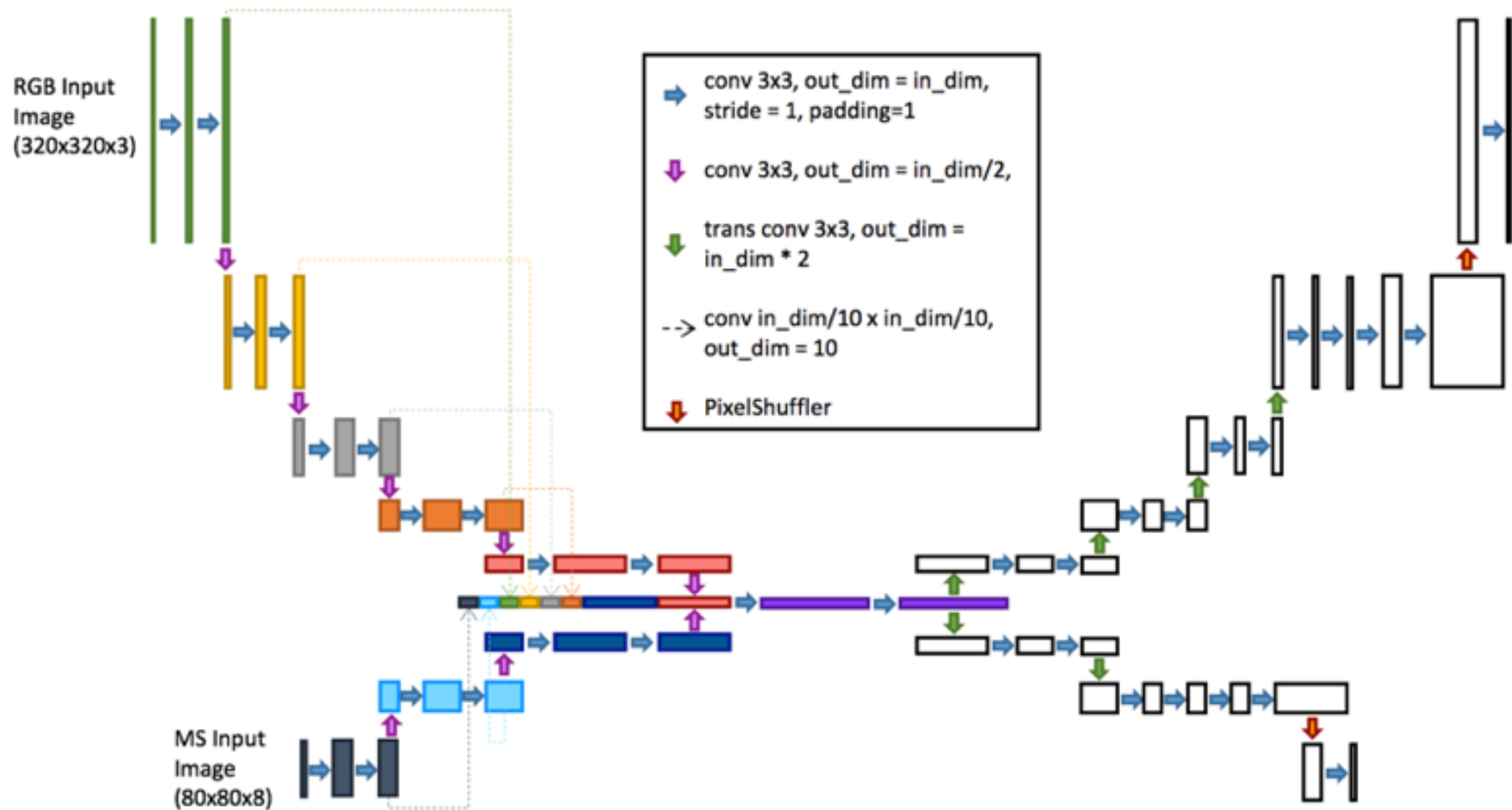
# Method

- We propose a deep learning architecture inspired by recent successful architectures such as U-Net and DenseNet
- Take a semi-supervised approach by using an autoencoder structure

# Joint Representation Learning



# Network Architecture

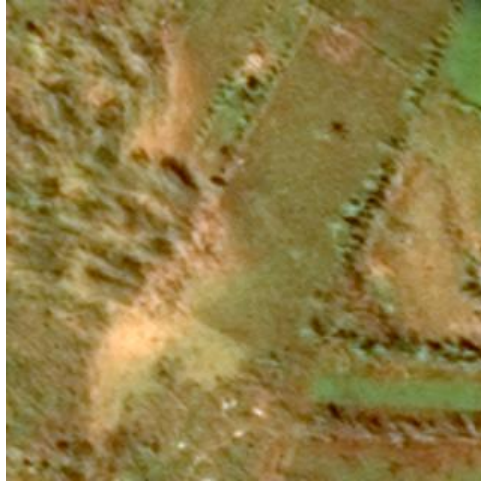


# Results

## Peak Signal to Noise Ratio

	Bimodal Autoencoder	Bidirectional DNN	X-Net (our architecture)
RGB	28.086	27.947	30.384
Multispectral	28.799	28.294	33.820

# Reconstruction Examples (RGB)



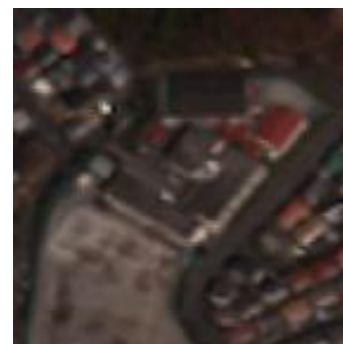
Input



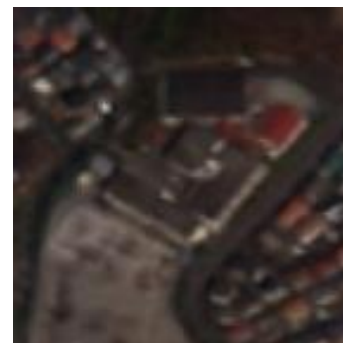
Reconstruction



# Reconstruction Examples (MS, RGB Channels Extracted)



Input



Reconstruction

# Future Work

- Establish an inference performance-based metric by evaluation tasks such as classification or change detection
- Explore variational autoencoder models and other analytical constraints to further improve the fusion layer
- Explore additional modalities and develop robustness methodologies to failing or missing modalities in training and testing

# References

- Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. arXiv preprint arXiv:1608.06993, 2016.
- Christie, G. et al (in press), “Functional Map of the World,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, (2018).
- Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. Multimodal deep learning. In International Conference on Machine Learning (ICML), Bellevue, USA, June 2011.
- O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *Proc. Int. Conf. Medical Image Comput. Comput.-Assisted Intervention*, pp. 234-241, 2015.
- V. Vukotić et al. Bidirectional Joint Representation Learning with Symmetrical Deep Neural Networks for Multimodal and Crossmodal Applications. In *Proceedings of the 2016 ACM International Conference on Multimedia Retrieval (ICMR)*, pages 343–346. ACM, New York, 2016.